HIV Population Dynamics in Vivo: Implications for Genetic Variation, Pathogenesis, and Therapy

John M. Coffin

Several recent reports indicate that the long, clinically latent phase that characterizes human immunodeficiency virus (HIV) infection of humans is not a period of viral inactivity, but an active process in which cells are being infected and dying at a high rate and in large numbers. These results lead to a simple steady-state model in which infection, cell death, and cell replacement are in balance, and imply that the unique feature of HIV is the extraordinarily large number of replication cycles that occur during infection of a single individual. This turnover drives both the pathogenic process and (even more than mutation rate) the development of genetic variation. This variation includes the inevitable and, in principle, predictable accumulation of mutations such as those conferring resistance to antiviral drugs whose presence before therapy must be considered in the design of therapeutic strategies.

Despite an extensive international research effort, HIV infection remains incurable and only modestly treatable. The infection of humans with this virus is characterized by three phases (1–4). Within several weeks after infection, there is an early phase with acute symptoms, extensive viremia, and large numbers of infected CD4-positive T cells in blood. Roughly coincident with the onset of the antiviral immune response (which includes antibodies and cytotoxic T cells), the amount of circulating virus declines by a factor of 100 or more (5, 6), leading to a clinically latent phase of variable duration with low but constant amounts of virus and infected cells in circulation and, usually, very gradually declining numbers of $CD4^+$ T cells. After about 10 years of clinical latency, the number of $CD4^+$ T cells declines to very low values and the symptoms of acquired immunode-

The author is in the Department of Molecular Biology and Microbiology, Tufts University School of Medicine, 136 Harrison Avenue, Boston MA 02111, USA.

ficiency syndrome (AIDS) appear, coincident with sharply increasing levels of virus and infected cells in circulation.

The ultimate collapse of the immune system despite the apparent low level of viral replication during the course of infection is the central paradox of AIDS. A resolution of this paradox may be at hand. Recent findings from a number of clinical trials now reveal HIV infection as an extraordinarily dynamic process in which a reasonably stable steady-state hides a surprisingly high rate of replication of virus and turnover of infected cells.

The problem addressed in these studies followed from a number of prior measurements of levels of virus or virus-infected cells in blood (*7–9*) or in lymph nodes (*10, 11*). Although a correlation between free virus and reduction in $CD4^+$ T cells could be obtained, it was impossible, from static measurements of this sort, to infer the dynamics of the virus-cell interactions underlying the steady-state values. To view the dynamics, it was necessary to perturb the steady state. This perturbation has been accomplished by blocking infection of cells in HIV-infected patients with lower than normal CD4 counts (*12*) by continuous treatment with potent inhibitors. The effect of the treatment was then measured by use of highly sensitive and reproducible assays for viral RNA in blood (*7, 13, 14*). Amazingly concordant results (summarized in Fig. 1) were obtained after treatment of HIV-infected patients with at least six different compounds, including nucleoside (*15, 16*) and nonnucleoside (*17–19*) inhibitors of reverse transcriptase (RT), as well as active-site inhibitors of the viral protease (*14, 17*). Prior to treatment, the concentration of virus in circulation remained essentially constant over periods of weeks to months. In all cases, however, there was a rapid decline in the concentration of circulating virus after treatment, with a half-time of about 2 days, and little variation from patient to patient. The concentration of virus declined to as little as 1% of the initial value at 1 to 2 weeks after treatment and then, in most patients, started to increase. When examined, the increase in viral RNA concentration was found to be due entirely to mutants resistant to the drug used (*15, 17, 20*). In contrast to the genome RNA pattern, analysis of proviral DNA in circulating cells revealed a high proportion of wild-type proviruses months after the start of treatment (*17, 21*). Finally, in all studies, there was a rapid increase in the number of CD4 cells in circulation, which declined back to near starting levels coincident with the rise in mutant genomes.

Inferences regarding HIV population dynamics in vivo rely on two principal assumptions. First, the treatments used must affect only new rounds of infection, not the lifetime of infected cells, the rate of production of virus from previously infected cells, or the clearance of free virus from the blood. This assumption is consistent with in vitro studies and is strongly supported by the consistency of data with different types of inhibitors, which are unlikely to share common side effects. Second, the virus population sampled (that in blood) must come from, or at least represent, the important replicating virus population, now believed to be predominantly in lymphoreticular tissue (*10, 22, 23*). It seems likely that there is relatively rapid (although perhaps inefficient) exchange of virions between the two compartments, but resolution of this issue will require more direct experimentation.

The conclusions of these studies lead to a model of the infection process that, although simple and somewhat incomplete, has novel implications for understanding important aspects of HIV pathogenesis. Included are the mechanism and significance of genetic variation of the virus, particularly development of drug-resistant mutants, and the role of virus replication and cell killing in the development of AIDS. Strategic considerations to guide the development of new approaches to therapy for HIV infection could follow directly from this improved understanding.

A Simple Steady-State Model

The significance of these experiments can best be understood in the context of a simple conceptual model of the infection process (Fig. 2). The very simplest model would include two populations of cells, one uninfected and one productively infected. The productively infected cells can be further divided into two phases: cells that are not yet virus-producing (DNA-positive, RNA-negative) and virus-producing cells (RNA-positive). At some point after the cells start to produce virus, they die (either by direct effect of the virus replication or as a result of a cytotoxic immune response). Newly infected cells are recruited by infection of the uninfected cell pool. If the system is at steady state, the rate of infection must equal the rate of death of infected cells, and the rate of replenishment of uninfected cells (by division or de novo generation) must equal the rate of loss of all cells combined. Also, the ratio of virus-producing to total infected cells will be equal to the average length of the productive phase divided by the lifetime of these cells. The rate of production of virus will be proportional to the number of virus-producing cells, and the steady-state concentration of virus in blood will be determined by the balance between its rate of production and rate of clearance. Finally, the average virus

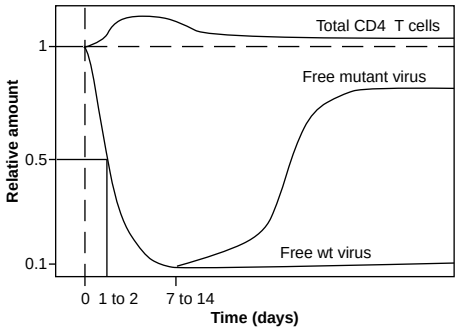

Fig. 1. Kinetics of virus load and cell number in HIV-infected individuals. The figure shows a composite of results from a number of studies monitoring the concentration in blood of wild-type and drug-resistant genomes by B-DNA or competitive reverse transcriptase–polymerase chain reaction (RT-PCR) assays as a function of time after drug treatment. The numbers shown on the abscissa indicate the range of values reported. The studies compiled here include trials on the nucleoside RT inhibitors 3TC (15) and AZT (16), nonnucleoside inhibitors nevaripine (17, 18), and delavirdine (19), and the protease inhibitors ABT-538 (14, 17) and L735,524 (17). The mutant virus curve is derived from results obtained with RT inhibitors; the status of virus reappearing after treatment with protease inhibitors has not yet been determined.

replication cycle time will be somewhat less than the average lifetime of the infected, virus-producing cells.

In addition to cells productively replicating virus, a few other types of infected cells must be added to the list. At the simplest, these include the following:

1) Latently infected cells, in which the nonproducing period is substantially longer than for the productive infection cycle discussed above. It has been proposed that some classes of nonactivated CD4 cells as well as immature monocytes may enter a period of latency after infection and not produce virus until activated (*24, 25*). Such cells, if present, would serve as a reservoir capable of maintaining the infection even after the most effective antiviral therapy. They could also provide a long-term memory of viral genotypes no longer present in replicating viral populations.

2) Chronically producing cells, which produce virus over an extended period of time, in contrast to the limited life-span of productively infected cells. With simple retroviruses (such as murine leukemia virus) there seems to be little death of infected cells, and chronically infected cells probably provide the major reservoir of virus. The existence of such cells in lentivirus infection is much less clear.

3) Defective proviruses. If the infection process involves frequent sequential infection of cells and the lifetime of infected cells is substantially less than that of uninfected cells, then the accumulation of cells

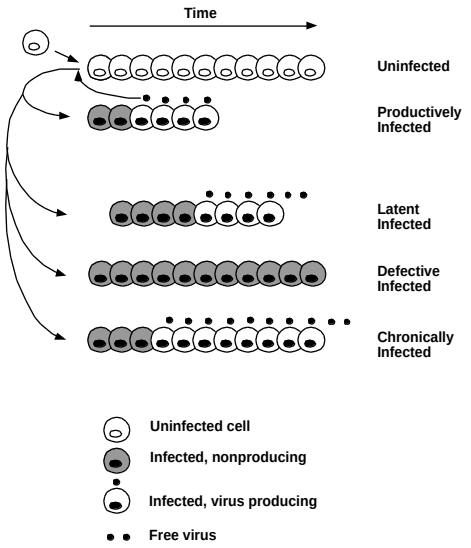

Fig. 2. A simple steady-state model for virus infection. Total cells are maintained at a constant level by replenishment (upper left) balancing the rate of death owing to natural causes as well as virus infection. Infection of cells moves them into one of three pools: productively infected cells with the most rapid turnover (second row), which proceed through a nonproducing phase followed by a producing phase and cell death; latently infected cells (third row), in which the nonproducing phase is considerably extended but (for the sake of example) the producing phase is identical to that of productively infected cells; cells with defective proviruses (fourth row), which do not produce virus and escape immune recognition; and chronically producing cells (fifth row), which produce virus for an extended period of time.

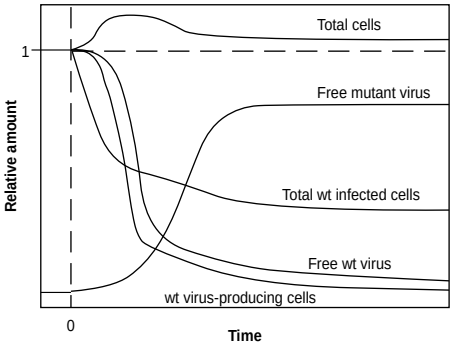

Fig. 3. Predicted effect on relative amounts of cells and virus after perturbation of the steady state at time 0 with a fully effective inhibitor of HIV infection.

containing proviruses with a variety of defects that prevent virus replication and cell killing is inevitable. Defects arise frequently as a consequence of errors during reverse transcription of wild-type genomes (*26*, *27*), and a fraction of these will have the property that the cell does not produce virus and does not die prematurely but preserves the defective provirus. Not all proviruses incapable of replication will have this property: many that express some viral proteins may still kill the infected cell or allow it to be recognized by the immune system. The survival of the cell containing a defective provirus would provide a powerful selection force for the provirus, since it would greatly outlive productively infected siblings.

Measurement of population sizes alone (including the numbers of infected and virus producing cells, the viral burden, and so on) can never give sufficient information to determine the relevant rates of production of viruses and cells or their lifetimes. For example, the static ratio of RNA-positive to DNA-positive cells, estimated at about 5:1 in lymph nodes of infected individuals (*11*), provides only a snapshot reflecting the ratios of the different infected cell types combined with the relative time each type of infected cell spends in the two phases. Determination of the actual times involved cannot be accomplished without perturbing the steady state.

If a population of infected cells in this steady state is treated with an inhibitor that rapidly, efficiently, and specifically blocks new rounds of virus infection, then the sequence of events diagrammed in Fig. 3 will occur. First, the total number of infected cells (as given by the quantitation of cells containing viral DNA) will begin to decline after a short lag representing the time of proviral DNA synthesis. The initial rate of decline will be equal to the average death rate (that is, the inverse of the mean lifetime) of all cell populations weighted by their size. This curve will decline through a series of slopes representing the turnover rates of the different infected cell populations (that is, productively infected, latently infected, and defective plus chronically infected). Given sufficiently accurate measurements, it should be possible to estimate the number, size, and lifetime of the different populations.

Second, after a lag reflecting the time between infection and the onset of cell death (the eclipse period), the number of cells producing wild-type virus will decline through a multiphase curve reflecting the death rates of the various virus-producing cell populations.

Finally, the amount of free virus will, after a similar lag, decline over time with a curve that reflects the balance between the rate of virus production and the rate of clearance or loss from the blood (by whatever means). Because the rate of virus production will be proportional to the number of virus-producing cells, the curve of declining virus levels will, of necessity, lag at least slightly behind the decline in virus-producing cells (*17*). Again, steps in the curve will reflect contributions to the virus populations from the various types of infected cells. In particular, the curve will level off at the end of the first phase at a relative amount of virus that places an upper bound on the contribution to the pool from latently and chronically infected cells.

Most importantly, although neither the death rate of infected cells nor the clearance rate of free virus can be estimated from the curve of free virus versus time, the curve does provide an upper bound for both these values at the same time. Thus, the midpoint of the curve must be greater than the mean lifetime of an infected cell and must also be greater than the half-life of circulating virus. Assume an initial steady state of circulating virus [that is, the rate of production (itself proportional to the number of virus-producing cells) equals the rate of clearance (proportional to the virus concentration)]. As the number of infected cells decreases as a result of drug treatment, the concentration of virus will decrease, but somewhat more slowly. The curve of virus concentration with time will most strongly reflect whichever is larger, cell death rate or virus clearance rate, but will always be at least slightly greater than either, and thus its midpoint provides a maximum estimate for both the mean lifetime of an infected cell and the half-time of virus clearance.

After the decline of wild-type virus, mutant virus resistant to the inhibitor used will increase in the cell population. The rate of increase is difficult to model, because it depends on the increase in the probability of transmission of the mutant virus to suitable uninfected cells in the absence of competition by wild-type viruses. There are, however, two important aspects of the kinetics of resistant mutants that can be understood. First, the timing of appearance of these mutants in the virus population will be strongly dependent on their frequency at the time of treatment. Second, their final steady-state level in the presence of the inhibitor should generally be less than that of wild-type virus before the initiation of treatment. Because the mutant virus will always replicate, at least slightly, more poorly than wild-type in the absence of inhibitor (otherwise, the mutation would be present in wild-type virus), it can be expected that it will also display a reduced replication capacity in its presence. This effect is likely to result in a steady-state level of infected cells reduced relative to what it would have been without treatment, and a corresponding reduction in the level of mutant virus found in blood.

As the number of productively infected cells declines, so does the rate of cell death, and a corresponding increase in the total number of cells ensues. Because replacement and cell death are initially in balance, the initial rate of increase in cell number will provide a lower bound on the cell death rate resulting from virus infection (*14*). As mutant virus reestablishes the infection and

the death rate increases, the cell number will decline to a value similar to (or slightly greater than) the initial value.

Application to the Infected Host

Although considerable work will be needed to refine the numbers, incisively test the assumptions, and formulate detailed kinetic models, the analyses described above already permit some important inferences based on the simple model described here.

First, the turnover of cells and virus is much greater than previously thought. A reduction of 100-fold in the level of virus within 2 weeks implies an average lifetime of an infected cell of less than 1 to 2 days, and the mean replication cycle time of the virus must be still less. Thus, the virus is replicating at a rate of some 300 or more cycles per year. In the median 10-year period of clinical latency, the virus genomes present in the infected individual are removed by 3000 generations from the virus that initiated the infection. On average, the virus that is transmitted to another individual will be over 1000 generations removed from the initial infection. This extent of replication per transmission cycle is probably without equal among viral (or possibly even bacterial) infections. Most other virus infections involve either short times (influenza and many others), truly latent infection of cells (as in herpesviruses), or a major role of chronically infected cells (as with other types of retroviruses).

Second, the decline of wild-type virus in circulation to 1% or less 1 to 2 weeks after the beginning of treatment shows that the contribution to the virus pool either from activation of latently infected or production from chronically producing cells is very small. More than 99% of the virus present at any one time is produced by cells infected within the previous week or two, and 30 to 50% from cells infected within the previous day. While a small population of latent or chronically infected cells may exist (as infected macrophages, for example), further experimentation based on assays with greater sensitivity and discrimination between mutant and wild-type genomes will be needed to quantitate them. Clearly, the majority of the infection is carried forward by repeated cycles of infection by virus from recently infected cells and not by infection at some earlier stage.

Third, the infection is directly causing a large turnover of target T cells. If the individual has a normal complement of about 2 x 10^{11} $CD4^+$ T cells (*14*), and 5% or more of these are productively infected (*23*), then substantially more than 10^9 cells die and must be replaced every day. A similar value can also be calculated from the initial rate of recovery of the $CD4^+$ T cell population after drug treatment (*14*). Although the normal replacement rate is not known, the rapid recovery of the $CD4^+$ T cells after the onset of treatment, followed by their decline as the virus load is reestablished, implies that the replenishment system is being severely stressed by the infection and that the depletion observed in CD4 cells is a direct result of their destruction, not a failure of their production (*14*).

Fourth, the presence of wild-type proviruses in circulating peripheral blood mononuclear cells (PBMCs) long after wild-type genomes have disappeared from the circulating virus population implies that a substantial fraction of these cells were infected months previously and are not representative of the majority of the replicating virus population. Indeed, it is highly likely that many or most of these cells contain defective proviruses selected by virtue of survival of the infected cell as discussed above. This conclusion is supported by the high frequency of defective genes (particularly *env* and *tat*) among sequences cloned from PBMC DNA (*28–32*). These considerations throw serious doubt on the common practices of interpreting the amount of proviral DNA in circulating PBMCs as a measure of "virus burden" and its genetic composition as a representation of the replicating virus pool.

Finally, the rapid and seemingly inevitable appearance and spread of resistant mutant virus after replication of wild-type virus is inhibited is consistent with the concept (discussed in more detail below) that these mutants are already present in the population at the time treatment is initiated. This conclusion is supported by the pattern in which distinct mutations appear that confer resistance to the nucleoside inhibitor 3TC (*15*).

These conclusions lead to a rather different model of the HIV infection process than has prevailed to date, which in turn has important implications for the understanding of HIV genetic variation, drug resistance, and mechanisms of pathogenesis.

Genetic Variation

The rapid replication and large population size of HIV in vivo imply that the virus can be considered as an ideal Darwinian population for the purposes of modeling genetic variation (*26*). Under conditions where there is a single population of virus undergoing large numbers of replication cycles without significant bottlenecks, and the population size is much larger than the inverse of the mutation rate, then the frequency of a mutation in the population can be estimated as a function of the frequency of its occurrence per cycle (that is, the mutation rate, μ) and its contribution to the fitness of the virus (the selective advantage, s). s is defined here as the relative growth rate (mutant/wild type) – 1. Thus, if the relative growth rate is 1.01, $s = 0.01$.

A simplified approach is possible because under these conditions all single (point) mutations will occur sufficiently often that they will be distributed over a background of all other mutations. Although the single-cycle mutation rate for HIV is not known, precedent from other retroviruses (*33–36*) suggests that it probably lies somewhere between 10^{-5} and 10^{-4} per base per cycle, with considerable variation from base to base depending on sequence context (*37, 38*). What is important is that the rate should be sufficiently high that the appearance of large numbers of mutations can be predicted. Given in vivo replication kinetics with more than 10^9 new cells infected every day, each and every possible single-point mutation occurs between 10^4 and 10^5 times per day in an HIV-infected individual.

Figure 4 shows a family of curves relating the frequency of a single point mutation to the number of replication cycles as a population arising from a single virus (as in transmission to a new host, for example). These curves were calculated simply by repeatedly incrementing the frequency at each cycle by the forward mutation rate and s (if positive) and decrementing it by the reverse mutation rate (assumed equal to the forward rate in all cases) and s (if negative).

This analysis shows that, after sufficient numbers of replication cycles, small positive values of s have a much stronger effect on the appearance of mutations (Fig. 4B) than large differences in mutation rate (Fig. 4A). Even values of s as low as 0.005 have a very strong influence on the accumulation of mutations over 1000 generations. Similarly, the accumulation of mutations that confer a selective disadvantage (those with negative values of s) is also very strongly influenced by s (Fig. 4C). Thus, the "evolution" of the HIV population into a set of related variants, or quasispecies (*39, 40*), does not reflect accumulation of neutral mutations driven by a high mutation rate, but rather the response of a sufficiently large and divergent population to a variety of subtle selection pressures. As I have argued (*26*), the influence of small selective forces in populations such as these is so strong that no mutation can be assumed to be truly neutral. It is clearly inappropriate to apply the principles and approaches of neutral theory (*41*) to such populations of viruses. In particular, the accumulation of mutations in such populations is a nonlinear function of the totality of selective forces acting on it. To calculate a "rate" at which mutations accumulate from observed diver-

gence in HIV and SIV (simian immunodeficiency virus) genomes and use this as a phenomenological clock to calibrate phylogenetic trees (*42*–*45*) is to invite errors in estimation of divergence times of many orders of magnitude.

That the extent of genetic diversity can be a strong function of the total replicative activity is emphasized by comparing the accumulation of mutations in the *env* genes of Rous sarcoma virus (a simple retrovirus from a genus that shows relatively little evolutionary diversity) during frequent cell culture passage (*26*, *46*) to that of cloned SIV replicating to monkeys (*47*) (Fig. 5). The similarity of the two curves implies that the difference in extent of diversity exhibited by these viruses in natural infections is not due to any fundamental difference in their underlying biology (such as a large difference in the mutation rate), but rather to the growth conditions of the virus population.

This pattern in which mutations accumulate also renders unlikely models of pathogenesis that rely on the appearance of some specific mutation [such as more cytopathic syncytium-inducing variants (*48*, *49*)] in HIV for the progression to late stage disease (*50*–*53*), because even extremely complex patterns of mutations will appear in the population quite early in the infection process. The failure of such viruses to become dominant in the population early in the infection must reflect their selective disadvantage at that time. By the same token, the appearance of such mutants late in disease must reflect the changing selective environment in the infected host, and although these viruses could still contribute to pathogenesis, they are more likely a symptom of end stage disease, than its cause.

Drug Resistance

To date, mutations with partial or complete resistance have been found for all compounds used or seriously considered as a therapy for HIV infection (*54*, *55*). Indeed, it is beginning to seem as though such mutations are an inevitable consequence of antiviral therapy. The pattern of disappearance of wild-type virus and appearance of mutant virus after treatment, combined with the coincident decline in CD4 cell numbers, strongly imply that, at least with some compounds, the appearance of mutant virus is the major cause underlying failure of the therapy.

The predicted kinetics for the appearance of slightly counterselective mutations, such as mutations to drug resistance in the absence of drug (Fig. 4C), implies that such mutations preexist in the virus population in substantial numbers at the time treatment is initiated. In the absence of inhibitor, these mutations must be at least slightly detrimental to virus replication (otherwise they would be present in wild-type virus). Because they will be generated at every round of replication, such mutants will initially accumulate at a rate equal to the mutation rate. After a sufficient number of cycles, however, they will approach a steady-state concentration of approximately $\mu/-s$, where s (a negative number in this case) is the selective disadvantage conferred by the mutation in the absence of inhibitor. (This value holds so long as $-s$ is much greater than the reverse mutation rate.) For example, if the forward and reverse mutation rates to resistance at a particular position are 10^{-4} per base pair cycle, and s is –0.01, then the steady-state frequency of this mutant in the population will be about $10^{-4}/-(-10^{-2})$ or 1% (*56*). Indeed, mutations known to confer resistance to a number of RT inhibitors have been seen at about this frequency in HIV proviruses in untreated infected patients (*57*).

Furthermore, so long as the effect of combined mutations on fitness of the virus is simply additive, then the frequency of a specific set of multiple mutations will be the product of the frequency of each one alone. Thus, many combinations of resistant mutations will also preexist in an infected individual, rendering unlikely the probability of eliminating the occurrence of mutation by completely blocking replication. A possible exception could occur in some cases where the quasispecies has not had time to develop, such as immediately after exposure to the virus.

The virtually inevitable and rapid occurrence of mutations with resistance to therapeutic agents has strong implications for drug design and application. Because the wild-type virus remains for only a short time after the onset of treatment, then no matter how effective a compound is at inhibiting replication of the virus, its clinical benefit is likely to be highly transitory. Nevertheless, the situation is not completely hopeless. Because drug-resistant mutants cannot replicate as well as wild type, it is to be expected that the steady-state level of mutant virus will be somewhat lower than that of wild-type, particularly if the mutation

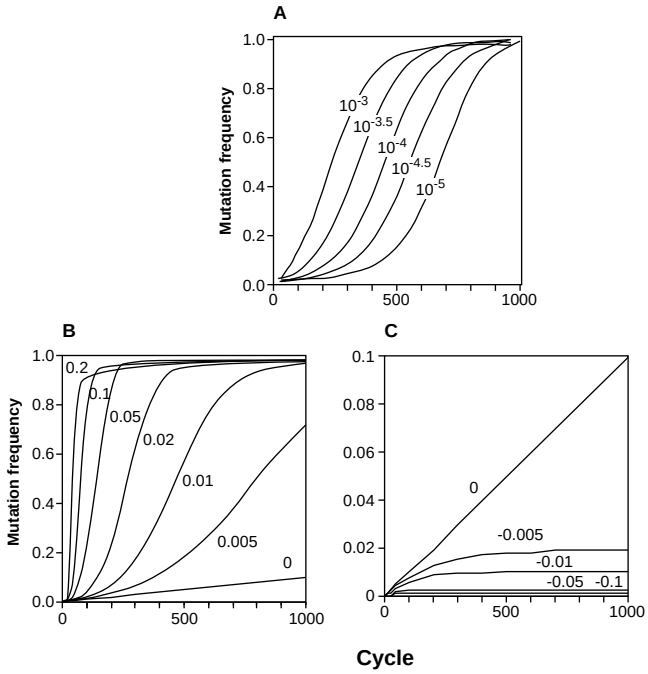

Fig. 4. Relation of mutation frequency to mutation rate, selective advantage, and number of replication cycles. The values were obtained with the use of a simple computer program (*71*) to "replicate" a virus population repeatedly, calculating the proportion of mutants in the population at each cycle from the forward and reverse mutation rates (μ) and the selective advantage (s) conferred by the mutation, starting with 0 mutants in the population. (**A**) shows the curves obtained for a variety of mutation rates as indicated, holding s constant at 0.01. (**B**) and (**C**) show the curves obtained at a variety of values of s, either positive (B) or negative (C) (note the expanded y axis), at a constant mutation rate (μ=10^{-4}). In all cases, the reverse mutation rate has been set as equal to the forward mutation rate. [Adapted from (*26*)]

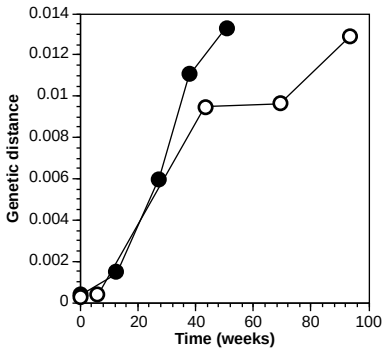

Fig. 5. Generation of diversity in complex and simple retroviruses in vivo and in vitro. The accumulation of point mutations in a region of the *env* gene obtained by PCR amplification of Rous sarcoma virus passaged repeatedly in cultures of chicken cells (●) (*46, 71*) is shown along with the similar values reported for a cloned SIV strain grown in rhesus monkeys (○) (*47*). In both cases, the total number of mutations in all sequences analyzed is plotted against time from the initial infection.

does not induce complete resistance. Even a partially reduced level of virus replication may be of considerable clinical benefit. It may also well be that there are drugs (or drug combinations) for which the replication disadvantage of the "best" combination of mutations is sufficiently great that the clinical latency will be extended almost indefinitely.

These considerations imply that a sound strategy for design of the next generation of antiviral therapies will include screening of promising compounds against all mutants that can be generated (initially in vitro, but a wider variety of mutants can be expected to exist in vivo). Compounds selected for further development will be those for which the "best" resistant mutant virus is still severely crippled, not necessarily those that exhibit the greatest effect on wild-type virus. It remains to be seen whether a compound for which only severely crippled resistant mutants arise in the short term can be good for many years of therapy. However, combinations of mutations that do not arise fairly soon after initiation of treatment must be severely crippled and are likely to take a very long time to appear. Such strategies will require lifelong treatment, because the wild-type virus will almost certainly take over the population after cessation of treatment and reestablish the preexisting infection. These considerations stress the importance of recognizing the inevitability of resistance mutations and incorporating this recognition into the drug development process.

Implications for Pathogenesis and Therapy

Even if the burden of circulating virus (and therefore, the number of productively infected cells) can be substantially and permanently reduced by a factor of 2 or 4 or 10, what would be the clinical benefit? Predicting the outcome depends in large part on the underlying relationship between virus replication and pathogenesis. The high rate of cell turnover raises the possibility that the stress on the CD4 T cell replacement system may alone be sufficient—after a median 10-year interval—to erode the body's ability to maintain an effective immune system (*14, 17*). This speculation is reinforced by the correlation between the amount of free virus (as RNA) in circulation and the reduction in CD4 cells (*7*). Thus, a 2-fold reduction in the number of productively infected cells might lead to a 2-fold increase in the mean clinical latency, and a 10-fold reduction might (in most cases) prolong latency indefinitely. Even the 2-fold effect would be highly significant clinically, especially by comparison to current therapy. It is entirely possible that some compounds and combinations of compounds currently under trial might well be capable of achieving a clinically useful, albeit less than complete, response (*14, 15*).

Such considerations may also provide a useful framework for understanding the results of studies on currently used therapies. For example, at least some of the mutations conferring resistance to azidothymidine (AZT) have very little effect on replication of the virus (that is, $-s$ is a very small number), as evidenced by the very slow reversion of these mutations to wild type during replication in the absence of inhibitor, both in vivo and in vitro (*58–60*), as well as the occasional appearance of such mutants in untreated individuals (*57*). In such a case, the effectiveness of the treatment could be of only a few months duration, until the original steady state is reestablished. If, indeed, immunodeficiency is due to accumulated damage over the course of the clinically latent phase, then treatment at almost any time during this period, or even after the onset of symptoms, could have a virtually identical effect in prolonging the time of progression to more serious disease or death. This effect could reconcile the apparent discrepancy between the results of the Concorde trial, in which no difference in rate of death or progression to more serious disease was seen in patients treated early or late after diagnosis of HIV infection (*61*), and earlier studies in which a significant difference in progression in AZT-treated as compared to untreated patients was observed in the short term (*62*).

The idea that HIV-induced immunodeficiency is due to direct killing of infected cells (either by the virus or by the immune system) is not highly popular at present, but clearly needs reconsideration. Many other pathogenic mechanisms have been proposed (*4*), including side effects of viral gene products, such as induction of apoptosis (*63–66*); inappropriate expression of cytokines by infected cells (*67, 68*), thereby disrupting immune signaling (*69, 70*); indirect effects on the architecture of the lymphoid tissue (*10*); and eventual evolution of variants that are more "virulent" or are better able to evade the immune response (*52*). The last mechanism seems much less likely in view of the rapid accumulation of diversity that must be occurring. At present, there is insufficient information to decide among the alternatives. Further in vivo experimentation involving perturbation of the HIV steady state could capture the correct information to help decide this important issue.

Regardless of the underlying mechanism of immunodeficiency, it is becoming more apparent that the engine that is driving the process is the constant repeated cycles of virus replication. Whether this will be true for other pathogenic effects of HIV, such as neurological disease or oncogenesis, remains to be seen. Phenomena such as virologic latency, chronic virus production by infected cells, and accumulation of defective variants are likely to be relatively unimportant to the overall process. It seems that the real disease caused by HIV occurs during the period when almost nothing seems to be happening, and that the final collapse of the immune system is the consequence of the accumulation of damage over the entire multiyear course of the infection, not the result of some new event happening after this lengthy interval. Irrespective of the mechanism, even a slight slowing of the process might be of considerable benefit to the infected patient.

Conclusions

Trials involving the use of effective inhibitors of HIV replication in vivo, combined with sensitive and accurate ways of measuring the quantity of wild-type and mutant genomes in circulation, have provided the first real look at the dynamic nature of HIV replication in vivo (*14–17, 19*). Application of even simple conceptual modeling reveals that the large majority of the infection process (at least as sampled in blood) comprises virus that is turning over more rapidly and in larger amounts than previously suspected. Indeed, it is this rapid turnover, not high mutation rate, cell tropism, or ability to establish proviral latency, that makes HIV (and other lentiviruses) unique among infectious agents. In the thousand or so replication cycles that separate infection of one individual from transmission to the next, subtle selective forces have an unparalleled opportunity to create the highly diverse quasispecies that characterize HIV populations. Constructive application of this line of reasoning to future experimentation could be highly rewarding in terms of revealing important properties of the virus-host interaction.

Future studies should elucidate the relation between the kinetics observed in blood and that in the cell populations at the sites of virus replication and dissect out the amounts and contributions of the various classes of infected cells at all stages of infection. Accurate assessment of viral clearance in blood will be necessary. It might be possible to do this by assessing relative levels of processed and unprocessed proteins in virions as a function of time after treatment with protease inhibitor. A more detailed analysis of the genetic consequences of drug treatment will be rewarding, such as an assessment of the preexisting levels of specific drug-resistant mutations in untreated individuals as well as the evolution of resis-

tant variants subsequent to treatment. Also, measurement of the genetic diversity of unrelated genes to the initial drug-resistant mutants as compared with the diversity of the wild-type virus will allow an assessment of the extent to which the treatment reduces the virus population and an independent estimate of the size of the preexisting mutant population. Finally, application of these principles to therapy will require determination of the clinical effect of modestly reduced levels of virus replication.

All of these studies are feasible with present-day technology and reagents. Their results will be of great value in finally coming to an understanding of the most hidden aspects of HIV pathogenesis and in applying this information to the development of effective therapeutic approaches. Such studies are possible only in patients in the course of trials involving potential therapeutic agents, and illustrate the great potential value of such clinical studies in illuminating fundamental scientific issues. Participants in these studies must be cared for first as patients, but at the same time their importance as a valuable and limited scientific resource must be considered. It is of the greatest importance that all such studies be designed to capture the basic information necessary to address the issues raised here.

In sum, the rapid appearance of HIV mutants resistant to even the most promising (and rationally designed) inhibitors may at first seem to be a discouraging turn of events for the development of effective anti-HIV therapies. In fact, the results obtained during the course of clinical trials are potentially so revealing that they may well light the way to both a real understanding of the infection process in vivo and to the development of truly effective therapeutic strategies involving new generations of inhibitors designed from the outset with the dynamics of the infectious process in mind.

REFERENCES AND NOTES

1. L. A. Evans and J. A. Levy, in *HIV Molecular Organization, Pathogenicity and Treatment*, W. J. W. Morrow and N. L. Haigwood, Eds. (Elsevier, Amsterdam, 1993), pp. 29-73.
2. G. Pantaleo, C. Graziosi, A. S. Fauci, *N. Engl. J. Med.* **328**, 327 (1993).
3. A. S. Fauci, *J. AIDS* **6**, 655 (1993).
4. J. A. Levy, *Microbiol. Rev.* **57**, 183 (1993).
5. E. S. Daar, T. Moudgil, R. D. Meyer, D. D. Ho, *N. Engl. J. Med.* **324**, 961 (1991).
6. S. J. Clark *et al.*, *ibid.*, p. 954.
7. M. J. Piatak *et al., Science*, **259**, 1749 (1993).
8. B. K. Patterson *et al., ibid.*, **260**, 976 (1993).
9. K. Saksela, C. Stevens, P. Rubinstein, D. Baltimore, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1104 (1994).
10. G. Pantaleo *et al.*, *Nature* **362**, 355 (1993).
11. J. Embretson *et al.*, *ibid.*, p. 359.
12. All the results contained so far were obtained from symptomatic patients with $CD4^+$ T cells >500 per milliliter. Strictly speaking, it is formally possible that the situation could be different in completely asymptomatic individuals, and this point should be experimentally addressed. However, the fact that the rate of decline of virus after treatment did not vary over a range of starting CD4 cell counts from <50 to about 500 (*14*) strongly suggests that the results will be similar at all stages of infection.
13. J. Mulder *et al.*, *J. Clin. Microbiol.* **32**, 292 (1994).
14. D. D. Ho *et al.*, *Nature* **373**, 123 (1995).
15. R. Schumann *et al.*, in preparation.
16. C. Loveday, personal communication.
17. X. Wei *et al.*, *Nature* **373**, 117 (1995).
18. D. D. Richman, personal communication.
19. W. G. Tarpley, personal communication.
20. D. D. Richman *et al.*, *J. Virol.* **68**, 1660 (1994).
21. M. J. Kozal, R. W. Shafer, M. A. Winters, D. A. Katzenstein, T. C. Merigan, *J. Infect. Dis.* **167**, 526 (1993).
22. G. Pantaleo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9839 (1991).
23. J. Embretson *et al., ibid.*, **90**, 357 (1993).
24. T. Folks *et al.*, *Science* **231**, 600 (1986).
25. K. Salsela, E. Muchmore, M. Girard, P. Fultz, D. Baltimore, *J. Virol.* **67**, 7423 (1993).
26. J. M. Coffin, *Curr. Top. Microbiol. Immunol.* **176**, 143 (1992).
27. V. K. Pathak and H. M. Temin, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6024 (1990).
28. S. Wain-Hobson, *Curr. Top. Microbiol. Immunol.* **176**, 181 (1992).
29. P. J. Johnson, T. E. Hamm, S. Goldstein, S. Kitov, V. M. Hirsch, *Virology* **185**, 217 (1991).
30. Y. Li et al., *J. Virol.* **65**, 3973 (1991).
31. A. Meyerhans *et al.*, *Cell* **58**, 901 (1989).
32. S. M. Wolinsky, personal communication.
33. V. K. Pathak and H. M. Temin, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6019 (1990).
34. R. J. Monk, F. G. Malik, D. Stokesberry, L. H. Evans, *J. Virol.* **66**, 3683 (1992).
35. J. P. Dougherty and H. M. Temin, *ibid.*, **62**, 2817 (1988).
36. A. Varela-Echavarria, N. Garvey, B. D. Preston, J. P. Dougherty, *J. Biol. Chem.* **267**, 24681 (1992).
37. M. Ricchetti and H. Buc, *EMBO J.* **9**, 1583 (1990).
38. G. J. Klarmann, C. A. Schauber, B. D. Preston, *J. Biol. Chem.* **268**, 9793 (1993).
39. M. Eigen and C. K. Biebricher, in *RNA Genetics*, E. Domingo, J. J. Holland, P. Ahlquist, Eds. (CRC Press, Boca Raton, FL, 1988), vol. 3, pp. 3–22.
40. J. J. Holland, J. C. De La Torre, D. A. Steinhauer, *Curr. Top. Microbiol. Immunol.* **176**, 1 (1992).
41. T. Gojobori, E. N. Moriyama, M. Kimura, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 10015 (1990).
42. T. F. Smith, A. Srinivasan, G. Schochetman, M. Marcus, G. Myers, *Nature* **333**, 573 (1988).
43. G. Myers and G. N. Pavlakis, in *The Retroviridae* J. Levy, Ed. (Plenum, New York, 1992), pp. 51-106.
44. R. F. Doolittle, D. F. Feng, M. A. McClure, M. S. Johnson, in *Retroviruses. Strategies of Replication*, R. Swanstrom and P. K. Vogt, Eds. (Springer-Verlag, New York, 1990), pp. 1-18.
45. M. Eigen and K. Neiselt-Struwe, *AIDS 1990* **4** (suppl. 1), 585 (1990).
46. J. M. Coffin, P. N. Tsichlis, C. S. Barker, S. Voynow, *Ann. N.Y. Acad. Sci.* **354**, 410 (1980).
47. D. P. W. Burns and R. C. Desrosier, *J. Virol.* **65**, 1843 (1991).
48. S. Schwartz, B. K. Felber, E.-M. Fenyo, G. N. Pavlakis, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7200 (1989).
49. E. M. Fenyo *et al.*, *J. Virol.* **62**, 4414 (1988).
50. H. Schuitemaker *et al.*, *ibid.*, **66**, 1354 (1992).
51. M. Tersmette *et al.*, *ibid.*, **63**, 2118 (1989).
52. M. A. Nowak *et al.*, *Science* **254**, 963 (1991).
53. C. Cheng-Mayer, D. Seto, M. Tateno, J. A. Levy, *ibid.*, **240**, 80 (1988).
54. R. Schinazi, B. Larder, J. Mellors, *Int. Antiviral News* **2**, 72 (1994).
55. C. Tantillo *et al.*, *J. Mol. Biol.* **243**, 369 (1994).
56. At each replication cycle, about 10^{-4} mutants will be created by new mutations and $0.01 \times 0.01 = 10^{-4}$ will be lost by negative selection.
57. I. Najera *et al.*, *J. Virol.* **69**, 23 (1995).
58. Q. Gao, Z. Gu, M. A. Parniak, X. Li, M. A. Wainberg, *ibid.*, **66**, 12 (1992).
59. D. D. Richman, *Curr. Top. Microbiol. Immunol.* **176**, 131 (1992).
60. H. Mohri, M. K. Singh, W. T. W. Ching, D. D. Ho, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 25 (1993).
61. Concorde Coordinating Committee, *Lancet* **343**, 871 (1994).
62. M. A. Fischl *et al.*, *N. Engl. J. Med.* **317**, 185 (1987).
63. C. Terai, R. S. Kornbluth, C. D. Pauza, D. D. Richman, D. A. Carson, *J. Clin. Invest.* **87**, 1710 (1991).
64. M.-L. Gougeon and L. Montagnier, *Science* **260**, 1269 (1993).
65. A. G. Laurent-Crawford et al., *Virology* **185**, 829 (1991).
66. H. Groux et al., *J. Exp. Med.* **175**, 331 (1992).
67. J. E. Merrill, Y. Koyanagi, I. S. Y. Chen, *J. Virol.* **63**, 4404 (1989).
68. J. E. Merrill, *Clin. Immunol. Immunopathol.* **64**, 23 (1992).
69. G. P. Linette, R. J. Hartzman, J. A. Ledbetter, C. H. June, *Science* **241**, 573 (1988).
70. B. Crise and J. K. Rose, *J. Virol.* **66**, 2296 (1992).
71. J. M. Coffin, C. S. Barker, D. Fiore, C. Naugle, in preparation.

72. I am indebted to S. Wolinsky for encouragement and repeated critical review, and to N. Rosenberg, R. Taplitz, D. Ho, A. Perelson, A. Neumann, G. Shaw, C. Boucher, and D. Richman for helpful suggestions. I would also like to thank D. Ho, G. Shaw, C. Boucher, C. Loveday, W. G. Tarpley, D. Richman, and their colleagues for generously sharing their unpublished data and ideas. Supported by grant R35 CA 44385 from the National Cancer Institute. J.M.C. is an American Cancer Society Research Professor. This paper is dedicated to the memory of Howard M. Temin.